



Universiteit
Leiden
The Netherlands

Can natural language processing provide accurate, automated reporting of wound infection requiring reoperation after lumbar discectomy?

Karhade, A.V.; Bongers, M.E.R.; Groot, O.Q.; Cha, T.D.; Doorly, T.P.; Fogel, H.A.; ... ; Schwab, J.H.

Citation

Karhade, A. V., Bongers, M. E. R., Groot, O. Q., Cha, T. D., Doorly, T. P., Fogel, H. A., ... Schwab, J. H. (2020). Can natural language processing provide accurate, automated reporting of wound infection requiring reoperation after lumbar discectomy? *The Spine Journal*, 20(10), 1602-1609. doi:10.1016/j.spinee.2020.02.021

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/4302490>

Note: To cite this publication please use the final published version (if applicable).

Clinical Study

Can natural language processing provide accurate, automated reporting of wound infection requiring reoperation after lumbar discectomy?

Aditya V. Karhade, BE^a, Michiel E.R. Bongers, MD^a, Olivier Q. Groot, MD^a,
Thomas D. Cha, MD, MBA^{a,b}, Terence P. Doorly, MD^c,
Harold A. Fogel, MD^{a,b}, Stuart H. Hershman, MD^a, Daniel G. Tobert, MD^{a,b},
Andrew J. Schoenfeld, MS, MSc^{d,e}, James D. Kang, MD^{d,e},
Mitchel B. Harris, MD^a, Christopher M. Bono, MD^a,
Joseph H. Schwab, MD, MS^{a,b,*}

^a Department of Orthopedic Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

^b Department of Orthopedic Surgery, Newton Wellesley Hospital, Newton, MA, USA

^c Department of Neurosurgery, North Shore Medical Center, Boston, MA, USA

^d Department of Orthopedic Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

^e Department of Orthopedic Surgery, Brigham and Women's Faulkner Hospital, Boston, MA, USA

Received 17 December 2019; revised 5 February 2020; accepted 28 February 2020

ABSTRACT

BACKGROUND: Surgical site infections are a major driver of morbidity and increased costs in the postoperative period after spine surgery. Current tools for surveillance of these adverse events rely on prospective clinical tracking, manual retrospective chart review, or administrative procedural and diagnosis codes.

PURPOSE: The purpose of this study was to develop natural language processing (NLP) algorithms for automated reporting of postoperative wound infection requiring reoperation after lumbar discectomy.

PATIENT SAMPLE: Adult patients undergoing discectomy at two academic and three community medical centers between January 1, 2000 and July 31, 2019 for lumbar disc herniation.

OUTCOME MEASURES: Reoperation for wound infection within 90 days after surgery

METHODS: Free-text notes of patients who underwent surgery from January 1, 2000 to December 31, 2015 were used for algorithm training. Free-text notes of patients who underwent surgery after January 1, 2016 were used for algorithm testing. Manual chart review was used to label which patients had reoperation for wound infection. An extreme gradient-boosting NLP algorithm was developed to detect reoperation for postoperative wound infection.

RESULTS: Overall, 5,860 patients were included in this study and 62 (1.1%) had a reoperation for wound infection. In patients who underwent surgery after January 1, 2016 (n=1,377), the NLP algorithm detected 15 of the 16 patients (sensitivity=0.94) who had reoperation for infection. In comparison, current procedural terminology and international classification of disease codes detected 12 of these 16 patients (sensitivity=0.75). At a threshold of 0.05, the NLP algorithm had positive predictive value of 0.83 and F1-score of 0.88.

CONCLUSION: Temporal validation of the algorithm developed in this study demonstrates a proof-of-concept application of NLP for automated reporting of adverse events after spine surgery.

FDA device/drug disclosure: not applicable.

Author disclosures: **AVK:** Nothing to disclose. **MERB:** Nothing to disclose. **OQG:** Nothing to disclose. **TDC:** Bio2 (B), GE Healthcare (B), Nuvasive (B), K2M (B). **TPD:** Nothing to disclose. **HAF:** Nothing to disclose. **SHH:** Nothing to disclose. **DGT:** Nothing to disclose. **AJS:** Nothing to disclose. **JDK:** Nothing to disclose. **MBH:** Nothing to disclose. **CMB:** Consulting:

United Health Care (B); Royalties: Elsevier (B), Wolters Kluwer (A); Other: The Spine Journal (NASS) (F). **JHS:** Consulting Stryker (B).

*Corresponding author. Joseph H. Schwab, MD, MS, Department of Orthopedic Surgery, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Boston, MA 02114.

E-mail address: jhschwab@mgh.harvard.edu (J.H. Schwab).

Adapting this methodology for other procedures and outcomes in spine and orthopedics has the potential to dramatically improve and automatize quality and safety reporting. © 2020 Published by Elsevier Inc.

Keywords: Adverse events; Artificial intelligence; Complication; Disc herniation; Infection; Machine learning; Natural language processing; Prediction; Reoperation; Spine

Introduction

Surgical site infections are a major driver of morbidity and increased health-care expenditures following spine surgery [1–3]. Value-based payment models and risk-based reimbursements make providers liable for postacute care. Prior studies have previously identified rates and risk factors for postoperative surgical site infections [4–7]. Others have also studied preoperative strategies to reduce these adverse events and developed predictive models for risk-stratification [3,8–10]. Despite the growing literature on characterization, prediction, and prevention of surgical site infections, there are few tools for automated surveillance of adverse events in spine surgery [11–13].

The current paradigm for surveillance of surgical site infections and other relevant clinical variables relies on prospective clinical tracking, manual retrospective chart review, or queries that use administrative procedural and diagnosis codes. These methods are labor intensive and have variable reliability [11,14–16]. Natural language processing (NLP) may be a promising alternative [11,17–19]. NLP is used for extracting structured information (occurrence of complications) from unstructured data (free-text, speech) [17]. The unstructured data are converted through a series of transformations into a representation that allows for content analysis and meaningful information extraction. For example, Murff et al. previously studied the utility of NLP for automated identification of 30-day postoperative acute renal failure, sepsis, deep vein thrombosis, pulmonary embolism, myocardial infarction, and pneumonia in patients undergoing major noncardiac surgery [11]. Rajkomar et al. examined the utility of NLP for automated clinical documentation from patient-physician conversations [19]. Liang et al. previously studied 570,000 pediatric patients and 1.4 million visits and applied NLP to automate the diagnosis of childhood diseases, achieving performance comparable to experienced pediatricians [20].

In this study, we sought to examine the utility of a machine learning-based NLP algorithm for automated surveillance of surgical site infections requiring reoperation in the 90 days following lumbar discectomy. The second aim of this study was to compare the performance of this NLP algorithm to Current Procedural Terminology (CPT) and International Classification of Disease (ICD) codes for surveillance of postoperative infections after lumbar discectomy. Lumbar discectomy was chosen in order to perform a “proof of concept” of this type of reporting.

Materials & methods

Guidelines

The transparent reporting of a multivariable model for individual prognosis or diagnosis guidelines were followed [21].

Data source

Our institutional review board approved retrospective review of medical charts at two academic medical centers and three community medical centers. Patients were identified through our retrospective institutional spine surgery registry. Manual chart review of electronic medical records was used to confirm the following inclusion and exclusion criteria for this study. The inclusion criteria were: (1) adult patients 18 years of age or older, (2) date of procedure between January 1, 2000 and July 31, 2019 (3) index procedure of lumbar discectomy for lumbar disc herniation. Patients who underwent decompression for any other operative indication or received other types of spine surgical procedures during the index surgery were excluded.

Outcome

The primary outcome was defined as reoperation for wound infection within 90 days of surgery. Free-text operative notes for all patients in the 90 days after surgery were reviewed by a clinical researcher (AVK) and discussed with senior authors. Operative notes for patients included in the testing set (surgery after January 1, 2016) were re-reviewed by two researchers (AVK, MERB) to confirm initial ratings. Discrepancies were adjudicated by mutual consensus with senior authors. Finally, wound cultures for all patients in the testing set were manually reviewed to confirm culture-proven wound infection.

Data analysis

Patients who underwent surgery before December 31, 2015 were used for algorithm training (n=4,483). Patients who underwent surgery after January 1, 2016 were used for algorithm testing (n=1,377). The following preprocessing steps were applied to raw free-text notes: (1) redundant white space and line breaks were removed (2) all reports were converted to lowercase (3) all punctuation was removed (4) stop words such as “the,” “and,” “with,” “who,” “for” were removed (5) all words in the free-text

notes were converted to their stemmed forms (ex. "infection," "infected," "infectious" converted to "infect"). An example of a raw and processed free-text note is available for review in [Supplementary Table 1](#). The bag-of-words method was applied to the processed text reports. In this method, a row is created for every free-text note and columns are created for words (tokens) in the operative notes. Entries in the matrix are the occurrence and frequency of words in the operative notes. Next, the results were regularized by using the term frequency-inverse document frequency technique to adjust for very rare and very common words/tokens.

An extreme gradient boosting (XGBoost) machine learning algorithm was trained on the free-text operative notes of patients who underwent surgery before December 31, 2015 to detect reoperation for infection within 90 days [22–24]. XGBoost is a machine learning algorithm that sequentially learns by preferentially weighting incorrect cases while using regularization techniques to prevent overfitting to the available data. The algorithm was assessed on temporal validation in the independent set of patients who underwent surgery after January 1, 2016. Metrics used for algorithm assessment were discrimination [area under the receiver operating curve (AUC-ROC), precision-recall curve (PRC), area under the precision-recall curve, sensitivity (recall), specificity, positive-predictive value (PPV, precision), negative-predictive value (NPV), F1-score (harmonic mean of precision and recall), positive likelihood ratio (LLR+), negative likelihood ratio (LLR-)], calibration (calibration intercept, calibration slope), and overall performance (Brier score) [25,26].

Next, the performance of a query combining procedural (CPT) and diagnostic (ICD codes, version 9 or version 10) for detection of reoperation for infection was assessed by the metrics described above. [Supplementary Table 2](#) lists the individual CPT and ICD codes.

Local and global explanations were generated for the NLP algorithm. Global explanations were provided through

relative variable importance plots to show the most important features (words) used for prediction. Local explanations were provided to show which features were used at the individual patient-level for prediction by highlighting the specific words used for prediction in free-text notes [27]. The Anaconda Distribution (Anaconda, Inc., Austin, Texas), R version (The R Foundation, Vienna, Austria), RStudio (RStudio, Boston, Massachusetts), and Python (Python Software Foundation, Wilmington, Delaware) were used for data analysis.

Results

Overall, 5,860 patients were included in this study and 62 (1.1%) had a reoperation for infection in the 90 days after surgery. The patients had a median age of 47 (interquartile range: 37–59) years and 2,628 (44.8%) were woman.

In the independent testing set of patients who underwent surgery after January 1, 2016 (n=1,377), the NLP algorithm achieved AUC-ROC of 0.999 ([Fig. 1A](#)), AUC-PRC of 0.96 ([Fig. 1B](#)), Brier score of 0.003, calibration intercept of 1.02, and calibration slope of 1.30 for detection of reoperation for infection [[Table 1](#)]. The AUC-ROC of the combined CPT and ICD codes was 0.88, AUC-PRC was 0.76, and Brier score was 0.003. In comparison, the null model Brier score (model estimating a probability equal to population prevalence of wound infection for every patient) was 0.011. In the testing set, the combination of CPT and ICD codes detected 12 of the 16 (sensitivity=0.75) patients who had reoperation for infection. However, the combination of CPT and ICD codes had perfect PPV and specificity [[Table 2](#)].

At a threshold of 0.05, the NLP algorithm detected 15 (true positives) of the 16 patients (sensitivity=0.94) who had reoperation for infection in the testing cohort and yielded three false positives. At this threshold, the algorithm had an F1-score of 0.88, specificity of 0.998, PPV of 0.83, and NPV of 0.999 in the independent testing set.

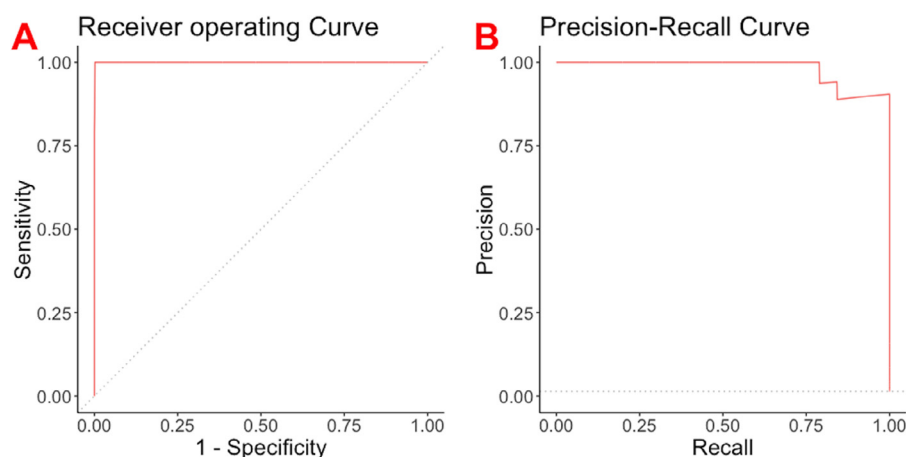


Fig. 1. (A) Receiver operating curve and (B) Precision-Recall curves for NLP algorithm on temporal validation in the independent testing set, n=1,377. NLP: natural language processing.

Table 1

Overall performance (95% confidence interval) of NLP algorithm for detection of postoperative infection relative to procedural and diagnosis codes in the independent testing set, n=1,377

	NLP algorithm	CPT and ICD codes
AUC-ROC	0.999 (0.99, 1.00)	0.88 (0.73, 0.96)
AUC-PRC	0.96 (0.84, 1.00)	0.76 (0.48, 0.94)
Brier	0.003 (0.001, 0.006)	0.003 (0.001, 0.006)
Calibration intercept	1.02 (0.06, 1.98)	-
Calibration slope	1.30 (0.86, 1.74)	-
Null model Brier score=0.011		

AUC-PRC, area under the precision-recall curve; AUC-ROC, area under the receiver operating curve; CPT, current procedural terminology; ICD, international classification of disease; NLP, natural language processing.

Decision curve analysis showed that changes based on the NLP algorithm yielded higher standardized net benefit compared to the default strategies of changes for all patients, changes for no patients, or changes based on the combination of CPT and ICD codes (Fig. 2).

Global relative variable importance plots for the NLP algorithm showed that the most important features for prediction of reoperation were ('infect', 'wound', 'debrid', 'seroma', 'place_prone', 'cultur', 'emerg', 'wound_infect', 'reveal', 'granul_tissu'). (Fig. 3). An example of a local explanation at the individual patient level is shown in Fig. 4; the algorithm used "wound," "infect," "cultur," "irrig," "drain" in the note to detect the occurrence of reoperation for wound infection.

Discussion

The rate of reoperation for infection in the 90 days after lumbar discectomy was 1.1% (n=62) in this study. This rate of infection is comparable to that of other large studies investigating infection after discectomy or lumbar decompression and reinforces the generalizability of our study findings [28–30]. On temporal validation among patients who underwent surgery on January 1, 2016 or later (n=1,377), the NLP algorithm detected 15 of the 16 patients

(sensitivity=0.94) who had reoperation for culture-proven wound infection. In comparison, the combination of CPT and ICD codes detected 12 of the 16 (sensitivity=0.75) patients who had reoperation for infection.

Within orthopedic surgery, others have previously used NLP for automated detection of periprosthetic femur fractures [31] and automated characterization of operative factors [32] among patients undergoing total hip arthroplasty. Within spine, NLP has been applied to automated detection of lumbar spine findings in radiology reports [33,34]. We previously developed a machine learning-based NLP algorithm for automated detection of intraoperative durotomy from free-text operative notes [13]. Despite the growing literature of NLP application within surgical disciplines, there are few studies that have used NLP for automated surveillance of postoperative adverse events after spine surgery [11,12].

Thirukumaran et al. previously developed rule-based NLP models to identify surgical site infections after orthopedic surgery [12]. For the rule-based NLP, the authors manually established 45 key words for infection based on a review of the literature and the clinical expertise of the study team. The occurrence and frequency of these a priori established key words (as well as multiword n-grams based on these keywords) in free-text notes were then used as input features for a logistic regression model. Advantages of this approach include verification of input features for the logistic regression model on the basis of consensus agreement of clinical experts. This leads to the avoidance of identification of non-clinically relevant features that may be present in the available data but have little to do with the primary outcome of the study. The authors used artificially balanced cohorts of 372 patients (172 patients with infection, 200 controls) during analysis. Further study of prospective patient populations may help delineate the sensitivity, PPV, and F1-score of these algorithms when the rates of infections more closely approximate clinical practice.

The current study applied a supervised machine learning-NLP approach whereby the features (words) were selected by the NLP algorithm from the training set without explicit programming. One advantage of the machine

Table 2

Performance (95% confidence interval) of NLP algorithm at various thresholds relative to procedural and diagnosis codes in the independent testing set, n=1,377

	CPT and ICD	NLP algorithm		
		Threshold=0.50	Threshold=0.10	Threshold=0.05
Sensitivity	0.75 (0.48, 0.93)	0.62 (0.35, 0.85)	0.88 (0.62, 0.98)	0.94 (0.70, 1.00)
Specificity	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	0.999 (0.99, 1.00)	0.998 (0.99, 1.00)
Negative predictive value	0.997 (0.99, 1.00)	0.996 (0.99, 1.00)	0.999 (0.99, 1.00)	0.999 (1.00, 1.00)
Positive predictive value	1.00 (0.74, 1.00)	1.00 (1.00, 1.00)	0.88 (0.62, 0.98)	0.83 (0.59, 0.96)
F1-score	0.86 (0.58, 0.96)	0.77 (0.47, 0.92)	0.88 (0.62, 0.98)	0.88 (0.64, 0.98)
LLR (+)	-	-	595 (147, 2,408)	425 (137, 1,326)
LLR (-)	0.25 (0.11, 0.58)	0.38 (0.20, 0.71)	0.13 (0.03, 0.46)	0.06 (0.01, 0.42)

CPT, current procedural terminology; ICD, international classification of disease; LLR-, negative likelihood ratio; LLR+, positive likelihood ratio; NLP, natural language processing.

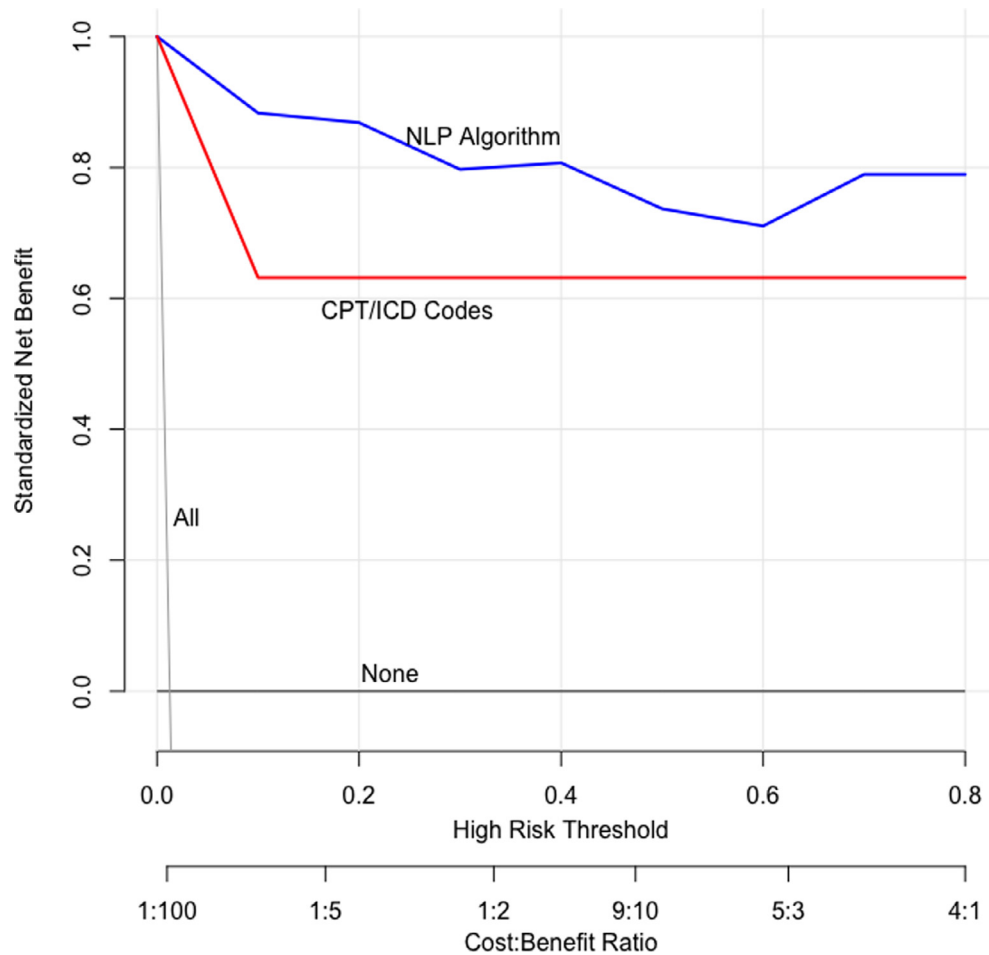


Fig. 2. Decision curve analysis with standardized net benefit of the NLP algorithm relative to default strategies of changes for no patients, or for all patients or changes based on the combined CPT and ICD codes. CPT, current procedural terminology; ICD, international classification of disease; NLP, natural language processing.

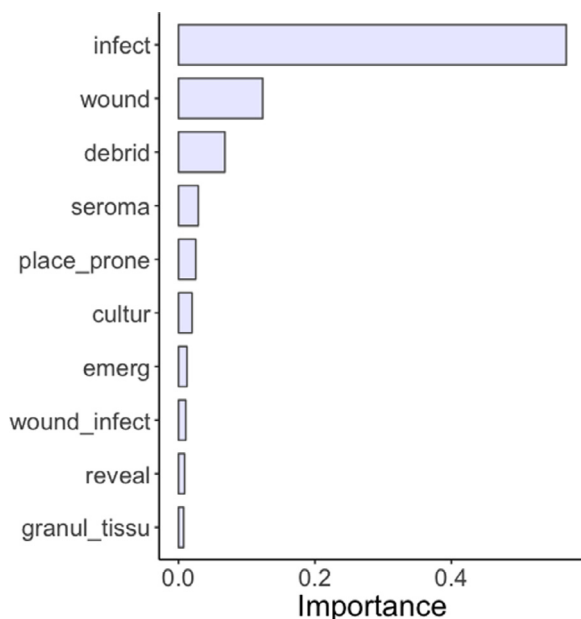


Fig. 3. Global explanation of features used by the NLP algorithm for prediction of reoperation for wound infection. NLP, natural language processing.

learning-based NLP approach relative to the rules-based approach may be more efficient generalization of this methodology to other procedures, indications, and outcomes on the basis of empirically observed trends in available data. This approach does not rely on explicit establishment of a new set of keywords for every new outcome or procedure of interest. One disadvantage of the machine learning-based NLP approach is the requirement of sufficient volumes of data for the algorithm to reliably abstract clinically meaningful terms without being explicitly programmed. Despite lack of explicit programming, the NLP algorithm in our study selected several features for prediction that have clinical relevance for postoperative wound infection (Fig. 3). However, the algorithm also identified features such as “place_prone” that have little clinical relevance for the primary outcome. One strength of the present study is the examination of the algorithm’s performance with temporal validation conditions where the data were not artificially balanced. Notably, even though only 16 of the 1,377 patients who underwent surgery after January 1, 2016 required reoperation for infection, the NLP algorithm showed good performance on sensitivity, PPV, and F1-score in identification of these patients.

preoper diagnosi post-op lumbar wound infect postop diagnosi post-op lumbar wound infect
 procedur incision drainag post-op superfici subfasci lumbar wound infect indic for procedur this
 39 year-old femal underw right-sid l5-s1 microdiscectomi she well oper discharg home post
 oper day 1 she present ed earlier even 2 day progress back right lower extrem pain mri
 lumbar spine demonstr contigu sub supra-fasci fluid collect concern infect per dr brown 's
 recommend fluid collect tap ultrasound guidanc reveal purul fluid as decis made proceed
 oper room urgent fashion lumbar wound explor washout discuss altern risk benefit oper
 patient descript of procedur the patient brought oper room intub place general anesthesia
 anesthesiolog periop antibiot held cultur obtain the patient posit prone jackson tabl pressur
 point appropri pad the lumbar region steril prep drape surgeri we began open prior incision
 cultur swab taken supra-fasci region we irrig copious amount bacitracin solut we dissect
 fascia use bovi cauteri we identifi previous place vicryl sutur proceed cut order explor sub-
 fasci space upon open fascia encount larg amount purul fluid swab aspir we irrig space
 copious amount bacitracin solut we inspect dural surfac not visual csf leak dural rent we
 proceed irrig entir caviti approxim 25 l bacitracin solut hemostasi obtain a hemovac drain
 place tunnel superior the incision close layer use vicryl deep layer run nylon sutur skin a steril
 dress appli the patient extub move extrem command she transfer pacu stabl condition

Label predicted: 1 (78.73%)

Explainer fit: 0.82

Fig. 4. Example of local explanation at the individual patient-level explanation for detection of reoperation for wound infection in a free-text operative note.

There are several limitations that must be acknowledged in order to appropriately interpret the results presented here. This was a retrospective analysis with the limitations inherent to study designs examining healthcare at single facilities. All of the medical centers included in this study are affiliated with one health-care entity. Some of the features, such as “place_prone”, used by the algorithm for detection of postoperative infections do not have direct clinical relevance to the primary outcome. This may represent overfitting of the algorithm to the available data. The true translational capacity of the algorithm developed in this study remains to be determined and external validation studies are required to support multi-institutional application. Retrospective review of free-text notes was conducted by two researchers and there is the potential for misclassification although we tried to minimize this to the fullest extent possible. Only wound infections requiring reoperation were included in this study and postoperative wound infections managed nonoperatively were not considered. Patients that underwent reoperation for wound infection at medical centers outside this health-care entity would also have not been captured. Event rates are fairly low, although well aligned with the risk of surgical site infection following discectomy as reported in the literature. This, combined with issues of patient clustering as noted above, may have restricted the clinical variation in cases of wound infection after surgery. The algorithm developed in this study was not designed to distinguish between types of wound infections (eg, epidural abscess vs superficial or subfascial

surgical site infection). Wound infection is relatively uncommon after lumbar discectomy and the generalization of this approach to more common postoperative complications of spine surgery or to other procedures with higher rates of postoperative infections remains to be determined. Reliable identification of postoperative infection requires appropriate documentation by surgeons. If clinicians fail to appropriately document postoperative infections in free-text notes, the algorithms presented here would be unable to detect the occurrence of these adverse events.

Despite these limitations, the algorithm developed here demonstrates a promising application of machine learning-based NLP for automated surveillance of adverse events after spine surgery. In this context, the analysis could be viewed as an exploratory proof of concept. The machine learning based-NLP approach described here could be adapted to other postoperative outcomes and procedures in the field of spine surgery. Current systems for surveillance of adverse events after spine surgery rely on manual chart review, dedicated prospective clinical documentation, or retrospective query based on administrative procedural and diagnosis codes. Machine learning based-NLP offers an alternative to these current approaches which all suffer from various inherent limitations. Well validated NLP algorithms have potential utility for hospital quality and safety reporting, national spine registries, and clinical outcomes research. In the future, NLP algorithms that continuously survey electronic health records for automated detection of adverse events may become integral to informatics-enabled spinal healthcare.

Conclusion

Temporal validation of the algorithm developed in this study demonstrates a proof-of-concept application of NLP for automated detection of adverse events after spine surgery. Adapting this methodology for other procedures and outcomes in spine may improve quality and safety reporting and clinical outcomes research.

Acknowledgment

Funding: The authors report no funding disclosures for this study.

Ethics statement: This study was approved by our institutional review board.

Supplementary materials

Supplementary material associated with this article can be found in the online version at <https://doi.org/10.1016/j.spinee.2020.02.021>.

References

- [1] Kang DG, Holekamp TF, Wagner SC, Lehman Jr RA. Intrathecal vancomycin powder for the prevention of surgical site infection in spine surgery: a systematic literature review. *Spine J* 2015;15:762–70.
- [2] Anderson PA, Savage JW, Vaccaro AR, Radcliff K, Arnold PM, Lawrence BD. Prevention of surgical site infection in spine surgery. *Neurosurgery* 2017;80(3s):S114–s23.
- [3] Tan T, Lee H, Huang MS, Rutges J, Marion TE, Matthew J. Prophylactic postoperative measures to minimize surgical site infections in spine surgery: systematic review and evidence summary. *Spine J* 2019.
- [4] Sebastian A, Huddleston 3rd P, Kakar S, Habermann E, Wagie A, Nassr A. Risk factors for surgical site infection after posterior cervical spine surgery: an analysis of 5,441 patients from the ACS NSQIP 2005–2012. *Spine J* 2016;16:504–9.
- [5] Cheng CW, Cizik AM, Dagal AHC, Lewis L, Lynch J, Bellabarba C. Body mass index and the risk of deep surgical site infection following posterior cervical instrumented fusion. *Spine J* 2019;19:602–9.
- [6] Karhade AV, Cote DJ, Larsen AM, Smith TR. Neurosurgical infection rates and risk factors: a national surgical quality improvement program analysis of 132,000 patients, 2006–2014. *World Neurosurg* 2017;97: 205–12.
- [7] Klemencsics I, Lazary A, Szoverfi Z, Bozsodi A, Eltes P, Varga PP. Risk factors for surgical site infection in elective routine degenerative lumbar surgeries. *Spine J* 2016;16:1377–83.
- [8] Lemans JVC, Oner FC, Wijdicks SPI, Ekkelenkamp MB, Vogely HC, Kruij MC. The efficacy of intrawound vancomycin powder and povidone-iodine irrigation to prevent surgical site infections in complex instrumented spine surgery. *Spine J* 2019;19:1648–56.
- [9] Janssen DMC, van Kuijk SMJ, d'Aumerie B, Willems P. A prediction model of surgical site infection after instrumented thoracolumbar spine surgery in adults. *Eur Spine J* 2019;28:775–82.
- [10] Dipaola CP, Saravanja DD, Boriani L, Zhang H, Boyd MC, Kwon BK. Postoperative infection treatment score for the spine (PITSS): construction and validation of a predictive model to define need for single versus multiple irrigation and debridement for spinal surgical site infection. *Spine J* 2012;12:218–30.
- [11] Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K. Automated identification of postoperative complications within an electronic medical record using natural language processing. *Jama* 2011;306:848–55.
- [12] Thirukumaran CP, Zaman A, Rubery PT, Calabria C, L Y, Ricciardi BF. Natural language processing for the identification of surgical site infections in orthopaedics. *J Bone Joint Surg Am* 2019.
- [13] Karhade AV, Bongers ME, Groot OQ, Kazarian ER, Cha TD, Fogel HA. Natural language processing for automated detection of incidental durotomy. *Spine J* 2019.
- [14] McLynn RP, Geddes BJ, Cui JJ, Ondeck NT, Bovonratwet P, Shultz BN. Inaccuracies in ICD coding for obesity would be expected to bias administrative database spine studies toward overestimating the impact of obesity on perioperative adverse outcomes. *Spine* 2018;43:526–32.
- [15] Nota SP, Braun Y, Ring D, Schwab JH. Incidence of surgical site infection after spine surgery: what is the impact of the definition of infection? *Clin Orthop* 2015;473:1612–9.
- [16] Quan H, Li B, Duncan Saunders L, Parsons GA, Nilsson CI, Alibhai A. Assessing validity of ICD–9–CM and ICD–10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res* 2008;43:1424–41.
- [17] Friedman C, Elhadad N. Natural language processing in health care and biomedicine. *Biomedical Informatics*. Springer; 2014. p. 255–84.
- [18] Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K. A guide to deep learning in healthcare. *Nat Med* 2019;25: 24–9.
- [19] Rajkomar A, Kannan A, Chen K, Vardoulakis L, Chou K, Cui C. Automatically charting symptoms from patient-physician conversations using machine learning. *JAMA Intern Med* 2019;179:836–8.
- [20] Liang H, Tsui BY, Ni H, Valentim CC, Baxter SL, Liu G. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med* 2019;25:433–8.
- [21] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015;13:1.
- [22] Xgboost: a scalable tree boosting system. In: Chen T, Guestrin C, editors. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM; 2016.
- [23] Torlay L, Perrone-Bertolotti M, Thomas E, Baci M. Machine learning—XGBoost analysis of language networks to classify patients with epilepsy. *Brain Inf* 2017;4:159.
- [24] Wang X, Zhang Y, Hao S, Zheng L, Liao J, Ye C. Prediction of the 1-year risk of incident lung cancer: prospective study using electronic health records from the state of Maine. *J Med Internet Res* 2019;21: e13260.
- [25] Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem* 2008;54:17–23.
- [26] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35:1925–31.
- [27] Why should I trust you?: Explaining the predictions of any classifier. In: Ribeiro MT, Singh S, Guestrin C, editors. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM; 2016.
- [28] Golinvaux NS, Bohl DD, Basques BA, Yacob A, Grauer JN. Comparison of the lumbar disc herniation patients randomized in SPORT to 6,846 discectomy patients from NSQIP: demographics, perioperative variables, and complications correlate well. *Spine J* 2015;15:685–91.
- [29] Rihn JA, Kurd M, Hilibrand AS, Lurie J, Zhao W, Albert T. The influence of obesity on the outcome of treatment of lumbar disc herniation: analysis of the Spine Patient Outcomes Research Trial (SPORT). *J Bone Joint Surg Am* 2013;95:1.
- [30] Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA* 2006;296:2441–50.
- [31] Tibbo ME, Wyles CC, Fu S, Sohn S, Lewallen DG, Berry DJ. Use of natural language processing tools to identify and classify periprosthetic femur fractures. *J Arthroplasty* 2019;34:2216–9.

- [32] Wyles CC, Tibbo ME, Fu S, Wang Y, Sohn S, Kremers WK. Use of natural language processing algorithms to identify common data elements in operative notes for total hip arthroplasty. *J Bone Joint Surg Am* 2019.
- [33] Tan WK, Hassanpour S, Heagerty PJ, Rundell SD, Suri P, Huhdanpaa HT. Comparison of natural language processing rules-based and machine-learning systems to identify lumbar spine imaging findings related to low back pain. *Acad Radiol* 2018;25:1422–32.
- [34] Huhdanpaa HT, Tan WK, Rundell SD, Suri P, Chokshi FH, Comstock BA. Using natural language processing of free-text radiology reports to identify type I modic endplate changes. *J Digit Imaging* 2018;31:84–90.